APPENDIX K

DRY WEATHER MODEL CONFIGURATION, CALIBRATION AND VALIDATION

The variable nature of bacteria sources during dry weather required an approach that relied on detailed analyses of flow and water quality monitoring data to identify and characterize sources. This TMDL used data collected from dry weather samples to develop empirical equations that represent water quantity and water quality associated with dry weather runoff from various land uses. For each monitoring station, a watershed was delineated and the land use was related to flow and bacteria concentrations. A statistical relationship was established between areas of each land use and flow and bacteria concentrations.

K.1 Background

Characterization of dry weather flow and indicator bacteria concentrations was based on analyses of data collected during studies of four watersheds in the San Diego Region. Two of these watersheds, Aliso Creek and San Juan Creek, are located in Orange County and are representative of conditions in the northern part of the Region (Figure 5-3). The remaining two watersheds, Rose Creek and Tecolote Creek, are located in San Diego County and discharge to Mission Bay (Figure 5-4). Three of these watersheds, Aliso Creek, San Juan Creek, and Tecolote Creek, are associated with water quality impairments due to bacteria and are therefore representative of conditions that may contribute to similar impairments in neighboring watersheds. Land uses for all four watersheds are consistent with other impaired watersheds in this study, with varying amounts of urban/residential land uses and open space in different subwatersheds.

To represent the linkage between source contributions and in-stream response, a mass balance spreadsheet model was developed to simulate source loadings and transport of bacteria in the impaired streams and streams flowing to impaired beaches. The model estimates bacterial concentrations to develop load allocations and to allow for future incorporation of new data. This predictive model represents the streams as a series of plug-flow reactors, with each reactor having a constant source of flow and bacteria. A plug-flow reactor can be thought of as an elongated rectangular basin with a constant level in which advection (unidirectional transport) dominates (Figure K-1).

The model segments are assumed to be well mixed laterally and vertically at a steady-state condition (constant flow and constant input). Variations in the longitudinal dimension are what determine any changes in parameters of concern. A "plug" of a conservative substance introduced at one end of the reactor will remain intact as it passes through the reactor. The initial concentration of bacteria can be entered for the injection point. At points farther downstream, the concentration can be estimated based on first- order die-off and mass balance.

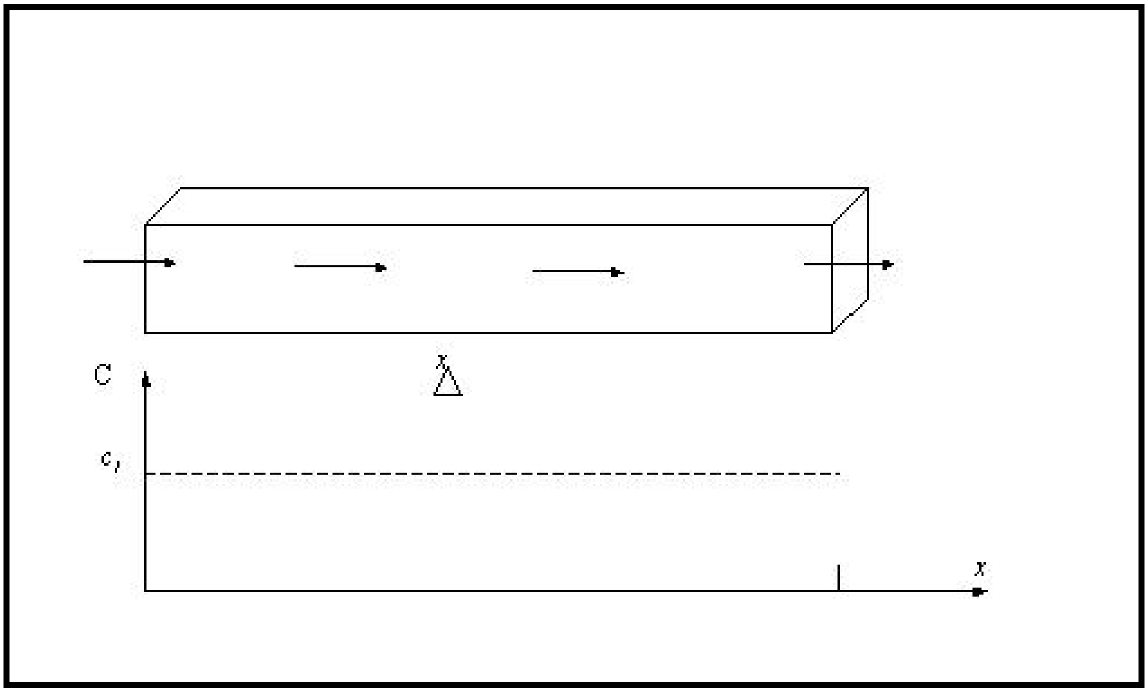


Figure K-1. Theoretical plug-flow reactor

This modeling approach relies on basic segment characteristics, which include flow, width and cross-sectional area. Model input for the flows and bacteria concentration of dry weather urban runoff was estimated using regression equations based on analyses of observed dry weather data. It is important to note that because each of these model parameters was estimated, the accuracy of the model is subject to the accuracy of the estimations. Bacteria concentrations in each reactor, or segment, are calculated using water quality data, a bacteria die-off rate, basic channel geometry and flow. Bacteria die-off rates, which can be attributed to solar radiation, temperature and other environmental conditions, were assumed first-order.

K.2 Model Configuration

Conceptually, the streams are segmented into a series of plug-flow reactors defined along the entire length of the stream to simulate the steady-state distribution of bacteria along its length. Multiple source contributions in a reactor are lumped and represented as a single input based on empirically derived inflows and bacteria concentrations. The model is one-dimensional (longitudinal) under a steady-state condition. Each reactor defines the mass balance for bacteria and water.

K.2.1 Physical Configuration

The first step in setting up and applying the model was the determination of an appropriate scale for analysis. Model subwatersheds were based on CALWTR 2.2 watersheds, stream networks, locations of flow and water quality monitoring stations, consistency of hydrologic factors and land use uniformity. The subwatersheds used in the dry weather model were the same as those used for the wet-weather model (see Appendix E for delineation of the subwatersheds).

Figure K-2 depicts an example of model connectivity of segments for the Chollas Creek watershed. Segments 1905, 1903, 1908 and 1907 are headwater segments. Segment 1902 begins where Segment 1903 and 1904 converge and so forth. For each model segment, mass balance is performed on all inflows from upstream segments, input from local watershed runoff, first-order bacteria die-off, stream infiltration and evaporation and outflow.

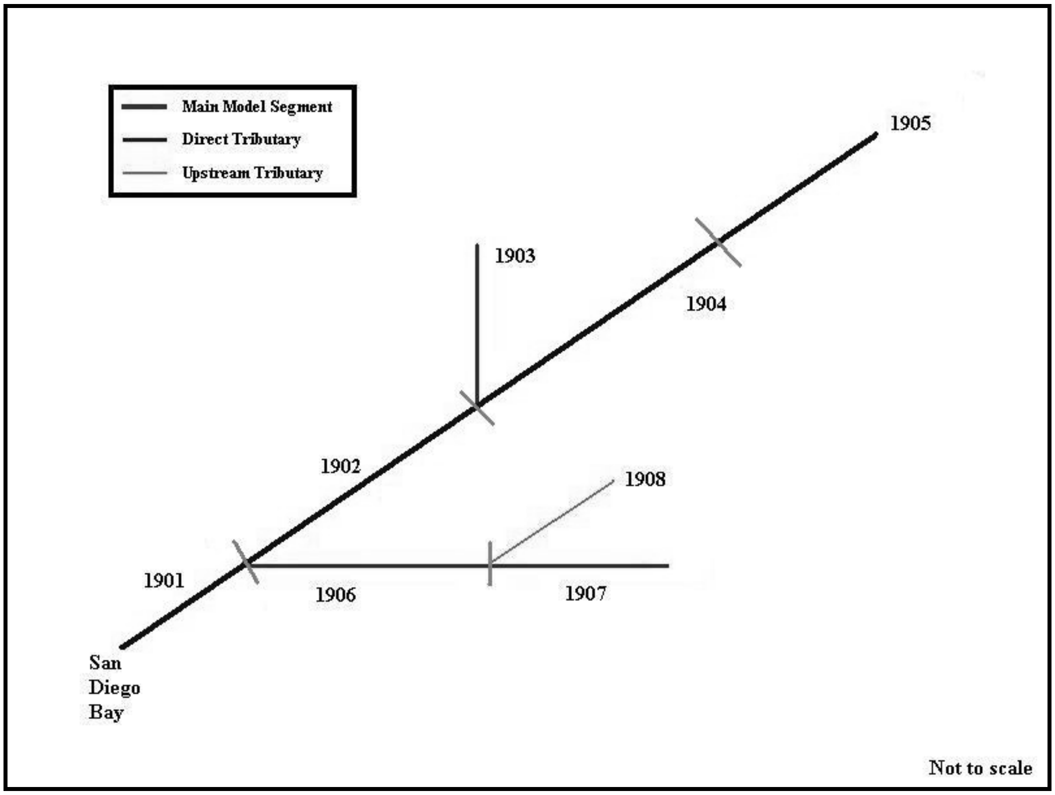


Figure K-2. Schematic of model segments for Chollas Creek and its tributaries

Using an upstream boundary condition of initial concentration (C_{in}) for inflow, the final water column concentration (C_{out}) in a segment can be calculated using the decay equation given below:

$$\frac{dc}{dt} = -kc \qquad \text{or} \qquad C_{out} = C_{in}e^{-kt} = C_{in}e^{-\left(k\frac{x}{u}\right)} \qquad (1)$$

where

C_{in} = initial concentration (#/100 mL)
C_{out} = final concentration (#/100 mL)
k = die-off rate (1/d)
χ = segment length (mi)
u = stream velocity (mi/d)

At each confluence, a mass balance of the watershed load and, if applicable, the load from the upstream tributary is performed to determine the initial concentration in the inflow to the reach. This is represented by the following equation:

$$C_{in} = \frac{Q_r C_r + Q_t C_t}{Q_r + Q_t} \qquad (2)$$

where

Q = flow (ft^3/s)
C = concentration (#/100 mL)

In the previous equation, Q_r and C_r refer to the flow and concentration from the receiving watershed and Q_t and C_t refer to the flow and concentration from the upstream tributary. The concentration calculated from this equation is then used as the initial concentration (C_{in}) in equation 1 for the receiving segment.

For calculation of outflows from the reach, the following equation is used. Infiltration rates for the model were determined through model calibration and comparison to literature ranges (see section K.5), and are dependent on stream length and width.

$$Q = Q_t + Q_r - I \qquad (3)$$

where

I = infiltration (ft^3/s)

Precise channel geometry data were not available for the modeled stream segments and therefore stream dimensions were estimated from analysis of observed data. Analysis was performed on streamflow data and associated stream dimension data from 53 USGS gages throughout Southern California. For this analysis, it was assumed that all streamflow at these gages less than 15 ft^3/s represented dry weather flow conditions. Using this dry weather data, the relationship between flow and cross-sectional area was estimated (R^2 = 0.51). The following is the resulting regression equation relating flow to cross-sectional area:

$$A = e^{0.2253 \times Q} \qquad (4)$$

where

A = cross-sectional area (ft^2)
Q = flow (ft^3/s)

In addition, data from the USGS gages were used to determine the width of each segment based on a regression between cross-sectional area and width. The best relationship (R^2 = 0.75) was based on the natural logarithms of each parameter. The following is the resulting regression equation from the analysis:

$$LN(W) = (0.6296 \times LN(A)) + 1.3003 \quad \text{or} \quad W = e^{((0.6296 \times LN(A)) + 1.3003)} \qquad (5)$$

where
W = width of model segment (ft)
A = cross-sectional area (ft^2)

K.3 Estimation of Dry weather Runoff

Flow data were not available for many of the subwatersheds. Estimates of inflows from the subwatersheds to the stream model were obtained through analysis of available data. Monitoring studies for which dry weather flow data were collected were available for Aliso Creek (performed by the Orange County Pubic Facilities and Resources Department and the Orange County Public Health Laboratory) and for Rose Creek and Tecolote Creek (performed by the City of San Diego) (Appendix G, No. 1 and 2). Information from these studies was assumed sufficient for use in characterizing dry weather flow conditions for the entire study area. For each study, flow data were collected throughout the year at stations throughout the watersheds. This information was used to understand the relationship between land use and stream flow.

An analysis was performed using dry weather data from the Aliso Creek (27 stations), Rose Creek (3 stations) and Tecolote Creek (2 stations) subwatersheds to determine whether there is a correlation between the respective land use types and the average of dry weather flow measurements collected at the mouth of each subwatershed. Table K-1 lists the stations and number of flow measurements used in this analysis. Selection of stations used in the analyses considered the number of flow measurements, the size of the watershed, as well as strategic locations of multiple watersheds representative of varied land uses. A linear relationship was established based on land use areas, with coefficients established through a *step-wise* multivariable regression analyses. For this regression, variables (land use areas) were added to the regression in a step-wise approach, and *p*-values were evaluated for each parameter. A *p*-value of less than 0.05 for each variable was used to determine their statistical significance. Some variables added at an early state of the regression analysis became statistically insignificant as additional variables were subsequently added to the model, which verified the necessity for a robust *step-wise* regression analyses over other more simplified methods. The resulting equation showed a good correlation between the flow and the commercial/institutional, open space and industrial/transportation land uses ($R^2 = 0.78$). The following is the resulting equation from the analysis (*p*-values for each variable are listed below):

$$Q = (A_{COM} \times 0.00168) + (A_{OPS} \times 0.000256) - (A_{IND} \times 0.00141) \qquad (6)$$

where
Q = flow (ft^3/s)
A_{COM} = area of commercial/institutional (acres) (p-value = 6E-13)
A_{OPS} = area of open space, including military operations (acres) (p-value = 0.029)
A_{IND} = area of industrial/transportation (acres) (p-value = 0.002)

Table K-1. Number of Flow Measurements at Each Station Used in Analyses

Watershed	Station	No. of Flow of Measurements
Aliso Creek	J01P08	35
	J01P06	21
	J07P02	40
	J07P01	38
	J01P01	40
	J01P05	39
	J01P03	40
	J01P04	40
	J06	15
	J05	39
	J01P30	39
	J01P28	39
	J01P27	40
	J01P33	40
	J01P25	40
	J01P26	40
	J01P24	35
	J01P23	40
	J01P22	39
	J03P02	39
	J01P21	32
	J02P05	39
	J02P08	40
	J03P13	38
	J03P05	40
	J03P01	39
	J04	6
Rose Creek	MBW11	7
	MBW13	80
	MBW16	76
Tecolote Creek	MBW7	23
	MBW9	77

Figure K-3 shows the predicted and observed flow data used in this regression.

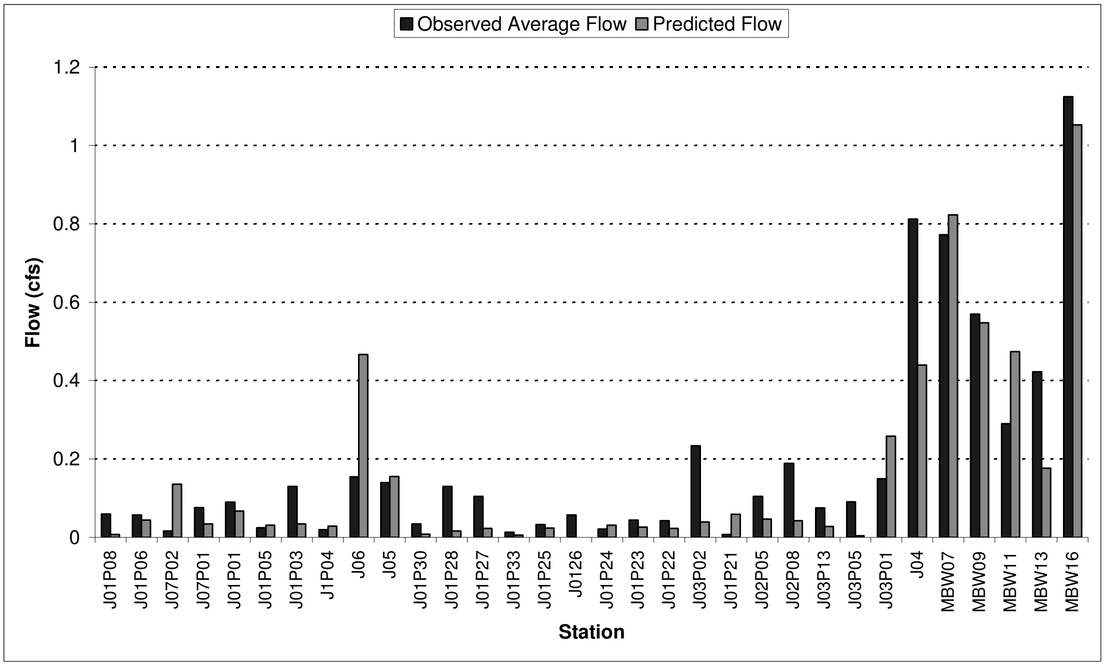


Figure K-3. Predicted and observed flows in Aliso Creek, Rose Creek and Tecolote Creek.

K.4 Estimation of Bacteria Densities

Like flow data, bacteria data were not available for many watersheds modeled. However, bacteria data had been collected for Aliso Creek (Orange County Pubic Facilities and Resources Department), San Juan Creek (Orange County Pubic Facilities and Resources Department) and Rose Creek and Tecolote Creek in the Mission Bay area (City of San Diego) (Appendix G, No. 4-6). For each study, multiple bacteria samples were collected throughout the year at stations throughout the watersheds. For this study, the information was used to understand the relationship between land use and water quality.

An analysis was performed using data from subwatersheds tributary to Aliso Creek (27 stations), Tecolote Creek (5 stations), Rose Creek (4 stations) and San Juan Creek (9 stations) to determine the correlation between dry weather fecal coliform concentrations, land use distribution and the overall size of the subwatersheds. For comparison, geometric means were calculated for each station using all dry weather data collected. Large data sets are preferred to reduce random error and normalize observations at each site. For example, if a station has 40 dry weather samples, the geometric mean of bacteria concentrations can be used for that station with confidence that they are representative of the range of conditions that normally occur. Likewise, if a station has only two samples, there is less confidence. It was critical that the data are normalized as well as possible before regression analysis so that variability does not propagate error. However, no criteria were developed for selection of stations based on the number of samples for representative geometric mean calculations. Rather, station selection included qualitative evaluation for consideration in the analyses. Specific stations of Rose Creek, Tecolote Creek, and San Juan Creek were selected for analyses even though few samples were available at these locations for geometric mean calculations. These stations were selected based on multiple reasons, including the relatively low indicator bacteria concentrations observed (see Figure K-4), strategic locations of watersheds to provide an expanded spatial coverage for analyses, size of the watershed, or representation of key land uses.

Table K-2. Number of Water Quality Samples at Each Station Used in Analyses

Watershed	Station	Number of Samples		
		Fecal Coliform	Total Coliform	Enterococci
Aliso Creek	J01P08	40	40	40
	J01P06	39	39	39
	J07P02	40	40	40
	J07P01	40	40	40
	J01P01	40	40	40
	J01P05	40	40	40
	J01P03	40	40	40
	J01P04	40	40	40
	J06	40	40	40
	J05	40	40	40
	J01P30	40	40	40
	J01P28	40	40	40
	J01P27	40	40	40
	J01P33	40	40	40
	J01P25	40	40	40
	J01P26	40	40	40
	J01P24	40	40	40
	J01P23	40	40	40
	J01P22	40	40	40
	J03P02	40	40	40
	J01P21	33	33	33
	J02P05	40	40	40
	J02P08	40	40	40
	J03P13	40	40	40
	J03P05	40	40	40
	J03P01	40	40	40
	J04	40	40	40
Rose Creek	MBW13	55	80	60
	MBW15	22	78	26
	MBW16	18	76	21
	MBW24	3	7	3

Table K-2. Number of Water Quality Samples at Each Station Used in Analyses (Cont'd)

Watershed	Station	Number of Samples		
		Fecal Coliform	Total Coliform	Enterococci
Tecolote Creek	MBW6	5	70	8
	MBW7	6	23	11
	MBW8	5	27	15
	MBW9	20	77	25
	MBW10	40	88	54
San Juan Creek	SJ13	11	11	11
	SJ14	10	10	10
	SJ15	11	11	11
	SJ16	11	11	11
	SJ19	3	3	3
	SJ20	11	11	11
	SJ21	11	11	11
	SJ29	2	2	2
	SJ32	11	11	11

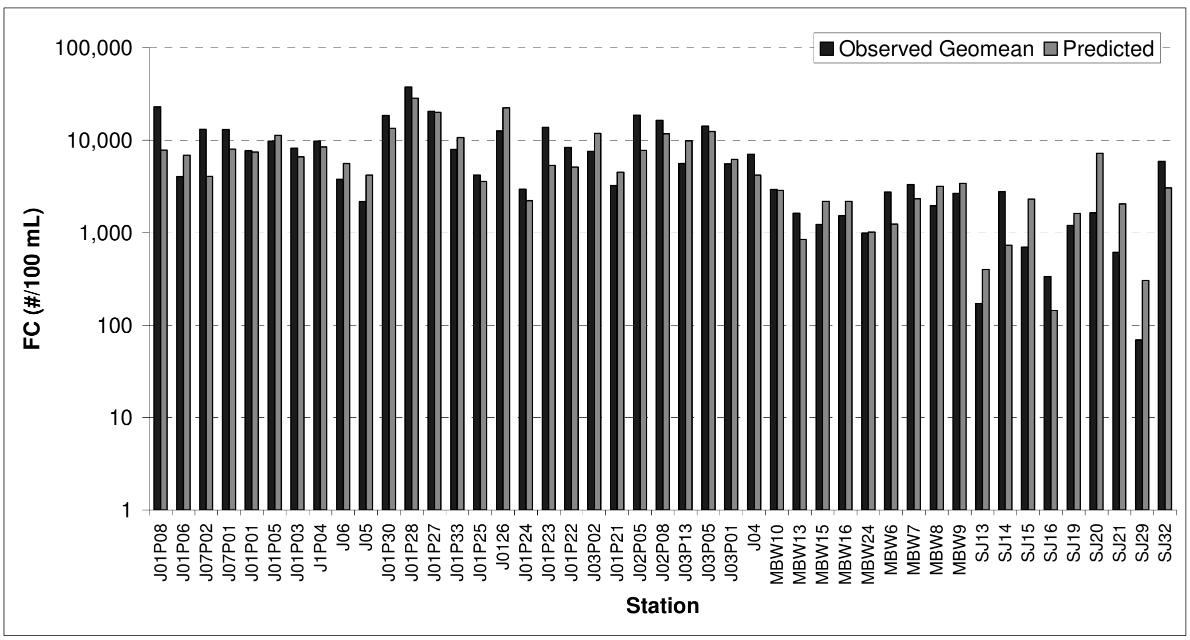
A regression analysis was then performed to determine whether there is a correlation between the representative geometric mean of fecal coliform data at each station, the percent of each land use category in the subwatershed and the total subwatershed area. Due to the variability of bacteria concentrations that often exceed multiple orders of magnitude, the analyses was based on the natural log of bacteria concentrations.

Coefficients in the equation were established through a *step-wise* multivariable regression analyses. For this regression, variables (percent of land uses) were added to the regression in a step-wise approach, and *p*-values were evaluated for each parameter. Percentages of land uses were used instead of land use areas since concentrations are not expected to increase with the size of the watershed, but rather due to the density of specific land uses. To include a function for reduction of bacteria concentration due to watershed size and increased potential for bacteria die-off (prior to entering the stream), an additional variable was added for watershed area. A *p*-value of less than 0.05 for each variable was used to determine their statistical significance (although this criterion was relaxed for open recreation which slightly exceeded at 0.067). As with the flow analysis, some variables added at an early state of the regression analysis became statistically insignificant as additional variables were subsequently added to the model, verifying the need for a robust *step-wise* regression analyses over other more simplified methods.

Results showed a good correlation between the natural log of fecal coliform concentrations and low-density residential, high-density residential, industrial/transportation, open space, transitional, commercial/institutional and recreation land uses, as well as subwatershed size (R^2=0.74). The following is the resulting regression equation from the analysis of fecal coliform concentrations (*p*-values for each variable are listed below). Figure K-4 shows observed geometric means and predicted concentrations to allow comparison.

$$LN(FC) = 8.48 \times (\%LU_{LDR}) + 9.81 \times (\%LU_{HDR}) + 8.30 \times (\%LU_{IND}) + 8.46 \times (\%LU_{OPS}) + 10.76 \times (\%LU_{TRN}) + 6.60 \times (\%LU_{COM}) + 17.92 \times (\%LU_{PRK}) + 12.85 \times (\%LU_{OPR}) - 0.000245 \times A \quad (7)$$

where: FC = fecal coliform concentration (#/100 mL)
$\%LU_{LDR}$ = percent of low density residential (p-value = 8E-16)
$\%LU_{HDR}$ = percent of high density residential (p-value = 7E-15)
$\%LU_{IND}$ = percent of industrial/transportation (p-value = 0.005)
$\%LU_{OPS}$ = percent of open space, including military operations (p-value = 7E-24)
$\%LU_{TRN}$ = percent of transitional space (p-value = 1E-19)
$\%LU_{COM}$ = percent of commercial/institutional (p-value = 4E-9)
$\%LU_{PRK}$ = percent of park/recreation (p-value = 0.009)
$\%LU_{OPR}$ = percent of open recreation (p-value = 0.067)
A = total area of watershed (acres) (p-value = 1E-7)



F

Figure K-4. Predicted versus observed fecal coliform concentrations.

The methodology for estimating fecal coliform concentrations was not as successful for prediction of total coliform and enterococci. Similar regression analyses were performed to determine whether there are relationships between total coliform and enterococci and land use and subwatershed size, but no acceptable correlations were found. As a result, a separate approach was used for estimating total coliform and enterococci concentrations in dry weather runoff for each subwatershed. For all stations in Aliso Creek, San Juan Creek, Rose Creek, and Tecolote Creek with five or more measurements of indicator bacteria concentrations (total of 170 stations), geometric means of fecal coliform, total coliform, and enterococci were calculated for each station and analyzed for trend analyses. This resulted in a single, normalized value of fecal coliform, total coliform, and enterococci at each station for comparison. Regression analyses were performed to determine whether there is a correlation between fecal coliform and levels of enterococci and total coliform. Results showed a good correlation for prediction of total coliform and enterococci as a function of fecal coliform (R^2=0.67 and R^2=0.77, respectively). The following are the resulting equations obtained (units of fecal coliform and total coliform/enterococci are consistent):

$$\text{total coliform} = 5.0324 \times \text{fecal coliform and}$$
$$\text{enterococci} = 0.8466 \times \text{fecal coliform} \qquad (8)$$

Figures K-5 and K-6 show comparisons of predicted (based on fecal coliform) and geometric means of observed total coliform and enterococci concentrations at each station.

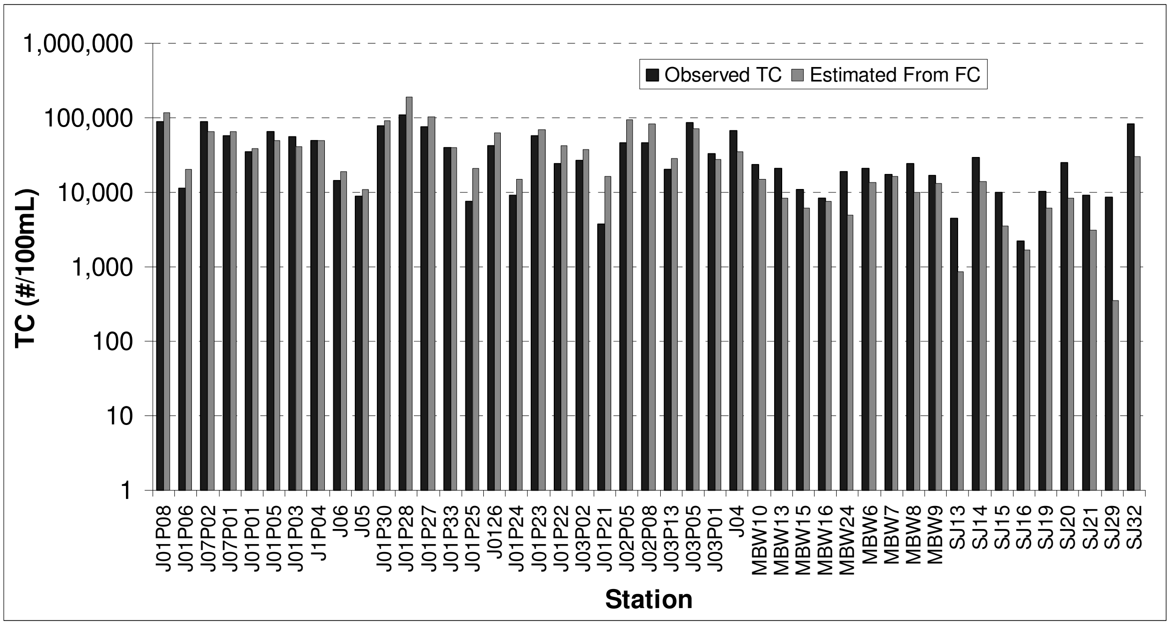


Figure K-5. Predicted versus observed total coliform densities

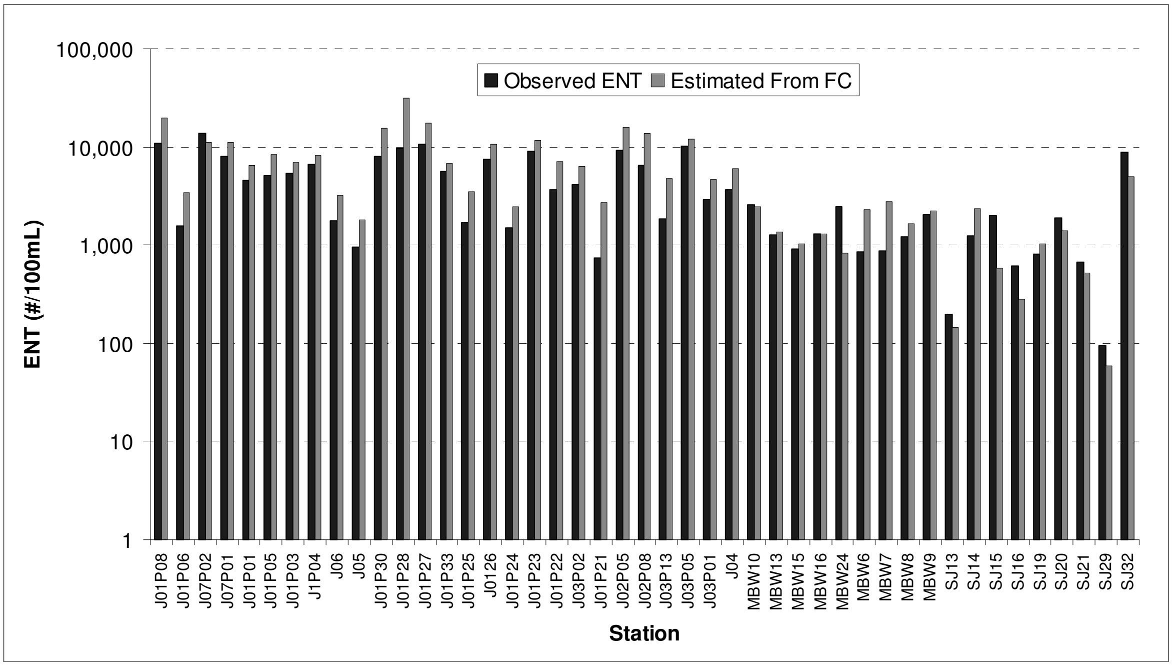


Figure K-6. Predicted versus observed enterococci densities

The above equations were used to estimate steady-state flows and indicator bacteria concentrations for each of the model subwatersheds. Several of the subwatersheds associated with monitoring stations used in the above analyses did not correspond to subwatersheds used in model development. For instance, stations on the Aliso Creek mainstem were used in regression analyses, and included the entire upstream watershed tributary to that location for characterization of land use and total area. However, model development of Aliso Creek included several smaller subwatersheds flowing into multiple segmented reaches that, although may result in a total watershed area consistent with the single watershed used in the regression analyses, differed in that stream infiltration and bacterial die-off rates in the multiple reaches must be defined. Therefore, model prediction of flows and bacterial concentration at locations on the Aliso Creek mainstem were based on upstream subwatershed loads predicted using the above equations, and routing through stream reaches that included assumptions for infiltration and bacterial die-off (based on model reach calibration and validation).

K.5 Model Calibration and Validation

Model assumptions for stream reach infiltration and bacterial die-off rates were derived through calibration based on data collected within reaches of Aliso Creek (11 stations) and Rose Creek (6 stations). Some of these stations were also used for development of regression equations for prediction of flow and fecal coliform concentrations from subwatersheds, however, effects of infiltration or bacteria die-off that may be implicitly incorporated in the regression equations (e.g., negative correlation of bacteria concentration to watershed size suggests effects of bacteria die-off in equation 7) were not considered duplicated in the reach assumptions. Model configuration of multiple subwatersheds and reaches differed from single representative watersheds used in regression analyses, and required incorporation of assumptions for reach infiltration and bacterial die-off to account for losses occurring during transport. Each model subwatershed used the regression equations to estimate flow and bacterial concentration that were routed through a network of stream reaches that ultimately met locations corresponding to monitoring stations used for calibration. However, watersheds used for regression analyses represented a single watershed for the same area, with no stream routing. Hence, the infiltration and die-off rates developed for the reaches were not consistent with errors associated with regression equations applied to the entire watershed without reach routing and losses considered. To further prove the independence of the calibration procedure from the regression analyses, data from five additional instream monitoring stations that were not used for regression analyses were also used for calibration. Model validation included nine additional stations not included in the regression analyses.

The calibration was completed by adjusting infiltration rates to reflect observed in-stream flow conditions and adjusting bacteria die-off rates to reflect observed in-stream bacteria concentrations. Following model calibration to in-stream flow and bacteria concentrations, a separate validation process was undertaken to verify the predictive capability of the model in other watersheds. Table K-3 lists the sampling locations used in calibration and validation, along with their corresponding watersheds. Figure K-7 shows the sampling locations in relation to the watersheds modeled for TMDL development (Appendix G, No. 4-6).

Table K-3. Calibration and Validation Sampling Locations

Calibration – Flow and Bacteria		Validation – Flow		Validation – Bacteria	
Watershed	**Sampling Location**	**Watershed**	**Sampling Location**	**Watershed**	**Sampling Location**
208	J01P22	403	USGS11047300	402	SJ04
209	J01P23	1701	MBW06	403	SJ05
210	J01P28	1702	MBW07	405	SJ18
211	J01P27	1703	MBW10	406	SJ24
212	J06	1704	MBW08	408	SJ1
213	J01P05	1705	MBW09	409	SJ29 & SJ17
214	J01P01			411	SJ06
215	J01TBN8			413	SJ08 & SJ07
219	J04			414	SJ30 & SJ09
220	J03P13			416	SJ15
221	J03P01			1701	MBW06
1601	MBW20			1702	MBW07
1602	MBW17			1703	MBW10
1603	MBW15			1704	MBW08
1605	MBW11			1705	MBW09
1606	MBW13				
1607	MBW24				

In the model, infiltration rates vary by soil type. Stream infiltration was calibrated by adjusting a single infiltration value, which was varied for each soil type by factors established from literature ranges (USEPA, 2000) of infiltration rates specific to each soil type. The goal of calibration was to minimize the difference between averages of observed streamflows and modeled flow at each station location (Figure K-7). Nine stations were used in calibrating the infiltration rate. The resulting infiltration rates were 1.368 in/hr (Soil Group A), 0.698 in/hr (Soil Group B), 0.209 in/hr (Soil Group C) and 0.084 in/hr (Soil Group D). The infiltration rates for Soil Groups B, C and D are within the infiltration range given in literature (Wanielisata et al., 1997). Soil Group A is below the range given in Wanielisata et al. (1997), however only one watershed in this TMDL is dominated by Soil Group A. Figure H-8 shows the results of the model calibration.

The modeled first-order die-off rate reflects the net effect on bacteria of various environmental conditions, such as solar radiation, temperature, dissolved oxygen, nutrients, regrowth, deposition, resuspension and toxins in the water. The die-off rates for fecal coliform, total coliform and enterococci were used as calibration parameters to minimize the difference between observed in-stream bacteria levels and model predictions. Calibration results for fecal coliform, total coliform and enterococci are presented in Figures K-9 through K-11. Die-off rates were determined for fecal coliform (0.137 1/d), total coliform (0.209 1/d) and enterococci (0.145 1/d). These values are within the range of die-off rates used in various modeling studies as reported by the USEPA (1985). Seventeen stations were used in calibrating die-off rates.

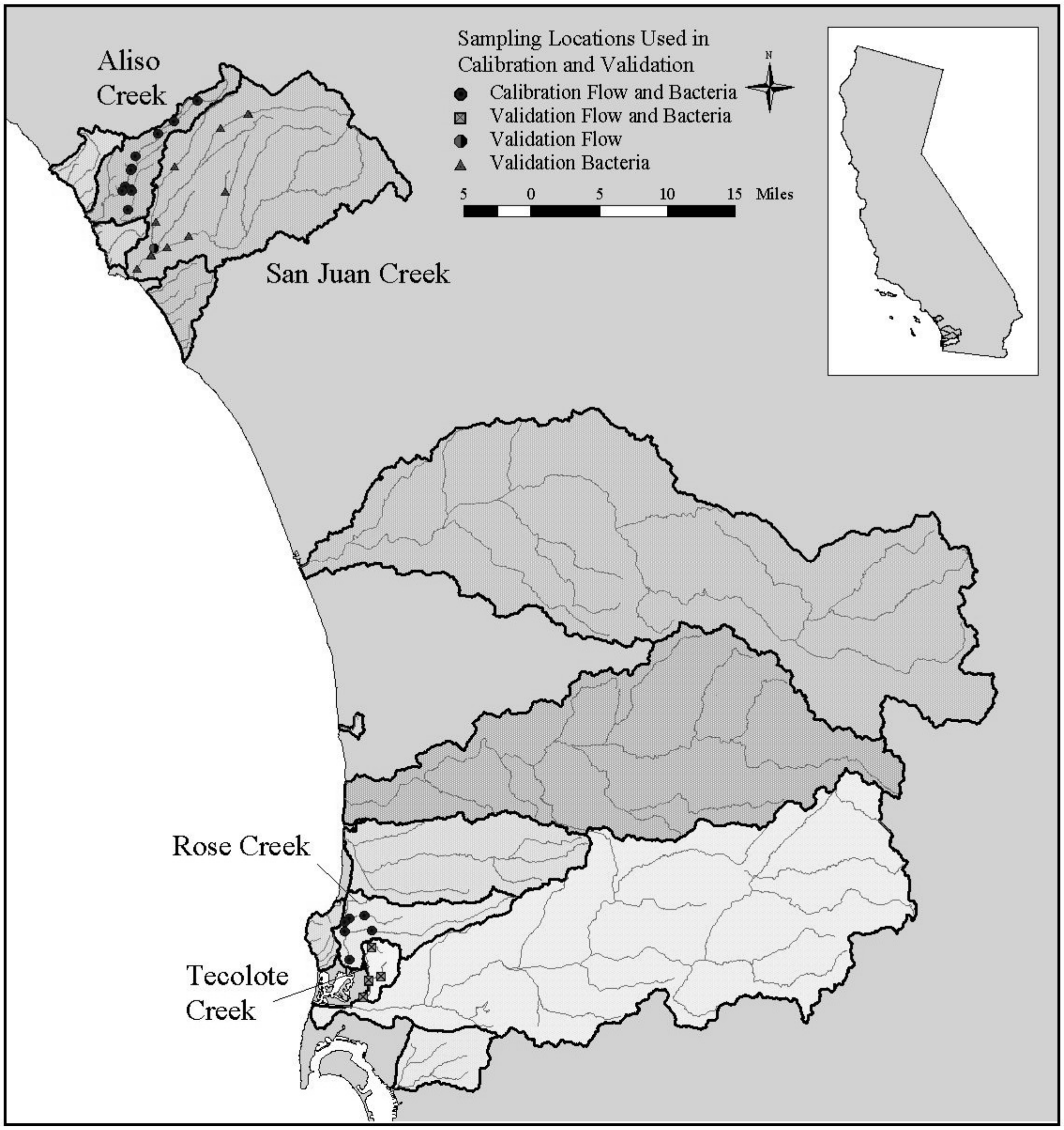


Figure K-7. Sampling locations used in model calibration and validation

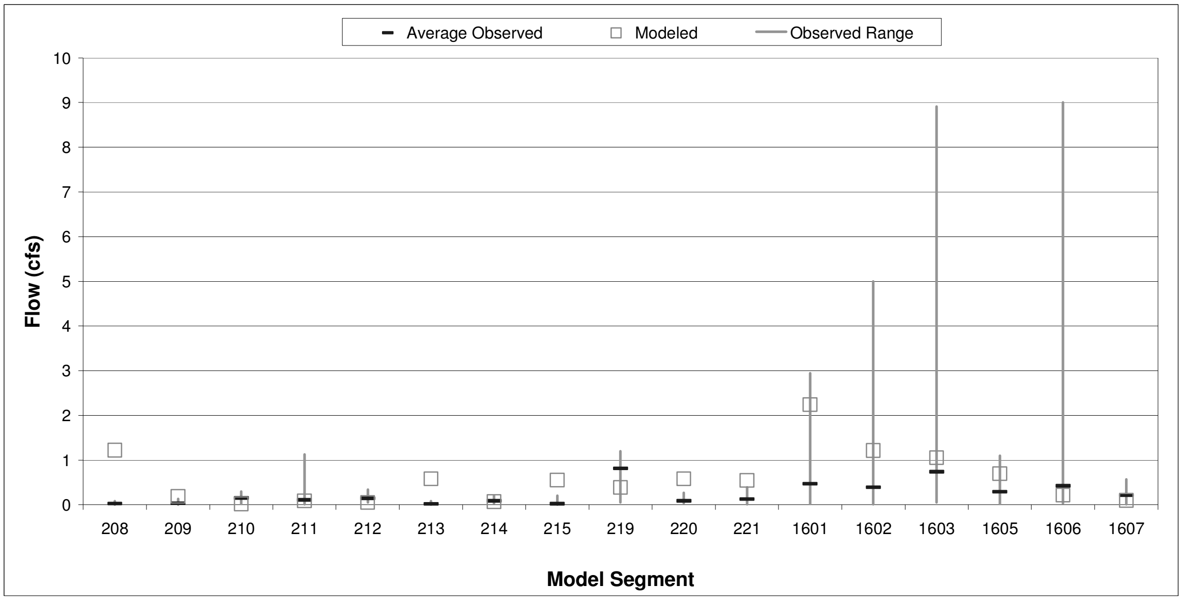


Figure K-8. Calibration modeled versus observed flows for Aliso Creek, Rose Creek and Tecolote Creek (Appendix B, No. 1 and 2)

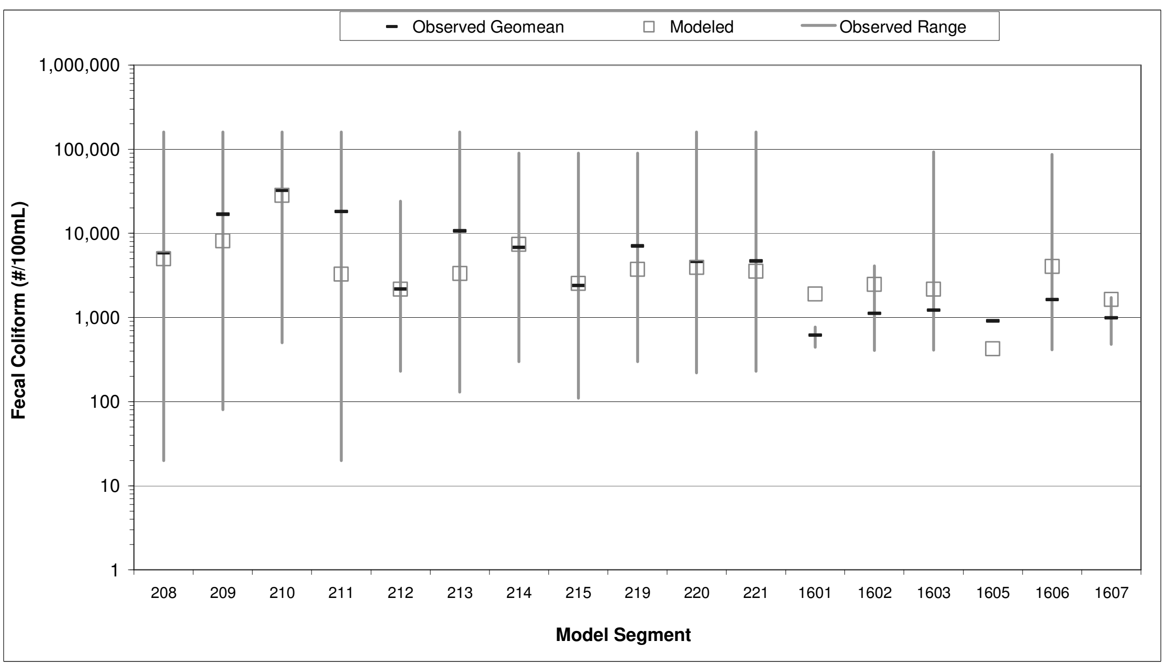


Figure K-9. Calibration modeled versus observed in-stream fecal coliform concentra-tions for Aliso Creek, Rose Creek and Tecolote Creek (Appendix B, No. 4 and 5)

.

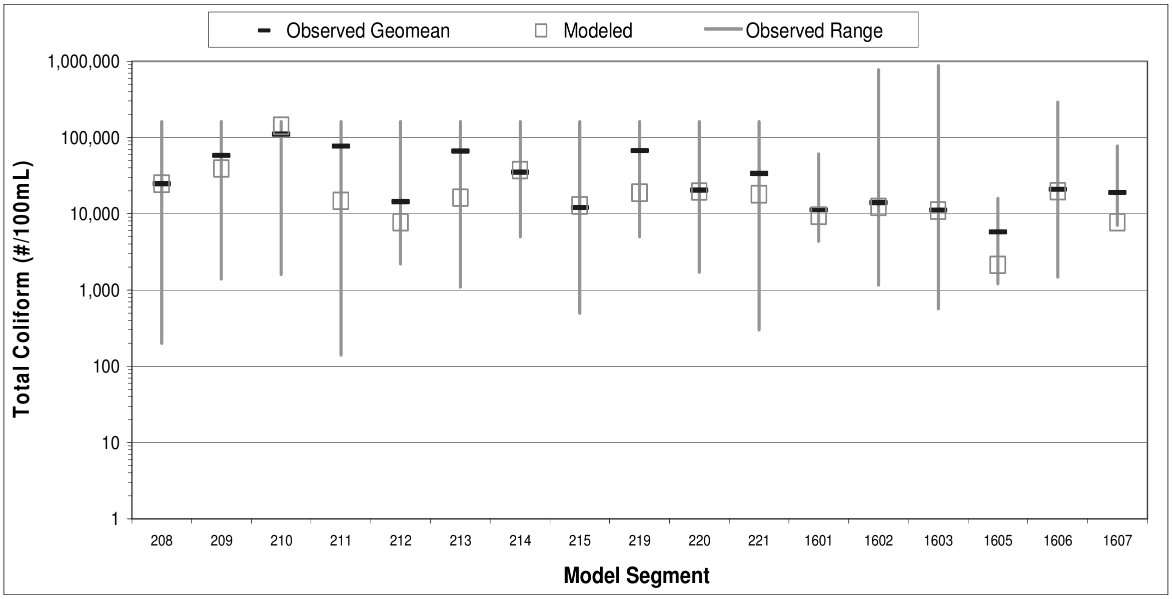


Figure K-10. Calibration modeled versus observed in-stream total coliform concentra-tions for Aliso Creek, Rose Creek and Tecolote Creek (Appendix B, No. 4 and 5)

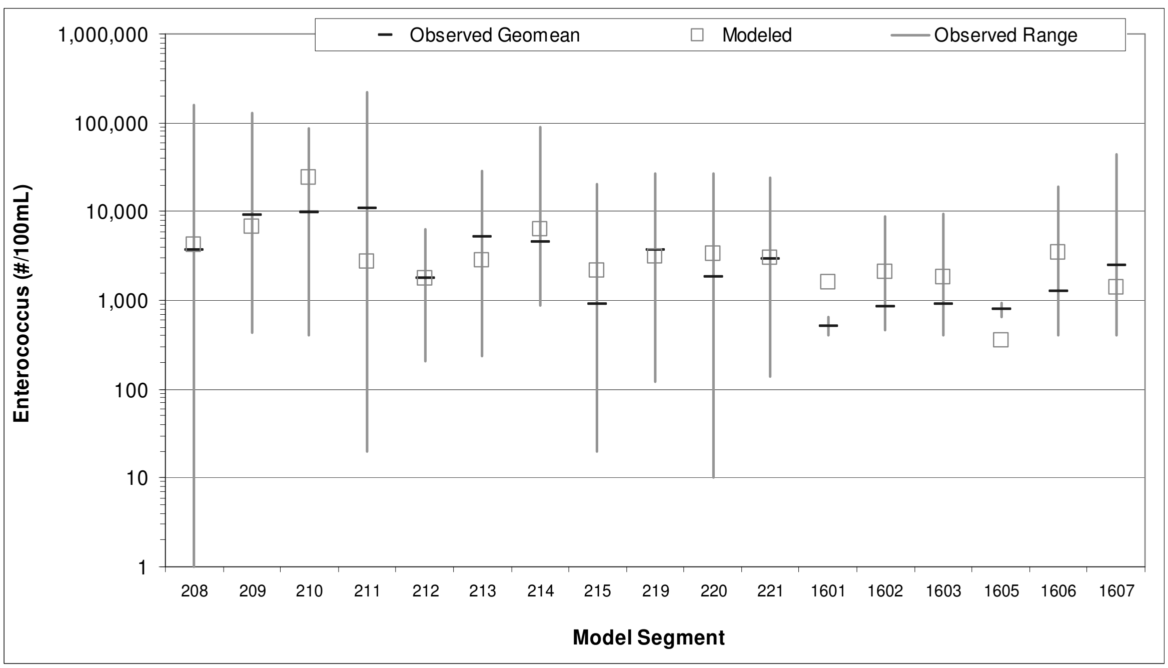


Figure K-11. Calibration modeled versus observed in-stream enterococci concentrations for Aliso Creek, Rose Creek and Tecolote Creek (Appendix B, No. 4 and 5)

The model was validated using six stations from San Juan Creek and Tecolote Creek (Appendix G, No. 2 and 3). One of these stations (USGS11047300) was not used in development of the regression equation 6. The model-predicted flows were within the observed ranges of dry weather flows (Figure K-12).

Model validation to in-stream water quality was provided using 15 stations on Tecolote Creek and San Juan Creek (Appendix G, No. 5 and 6). Eight of these stations were not used in development of the regression equation 7. The results of the water quality validation are presented in Figures K-13 through K-15.

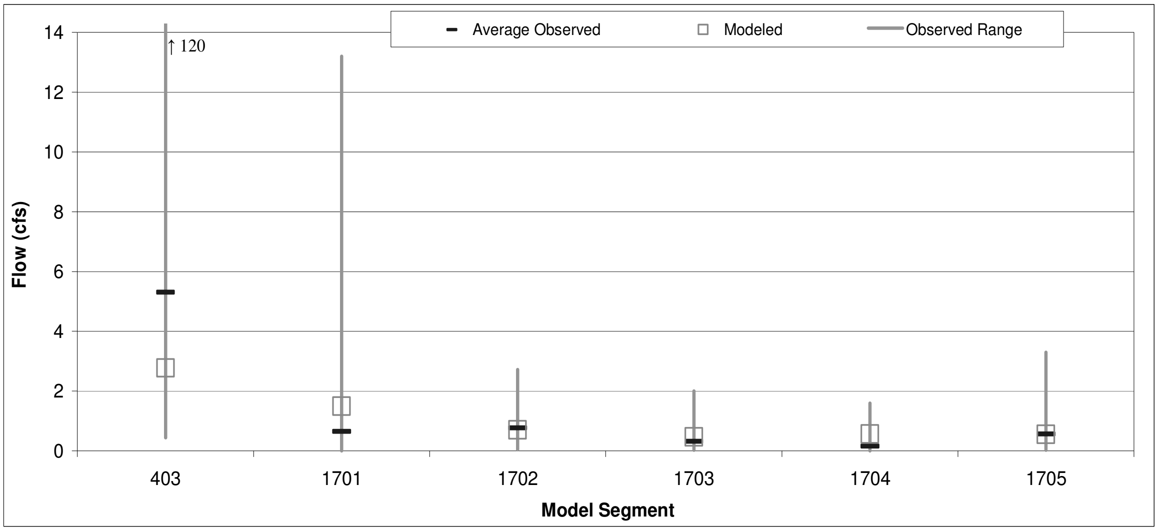


Figure K-12. Validation of modeled versus observed streamflow for San Juan Creek, Rose Creek and Tecolote Creek (Appendix B, No. 2 and 3)

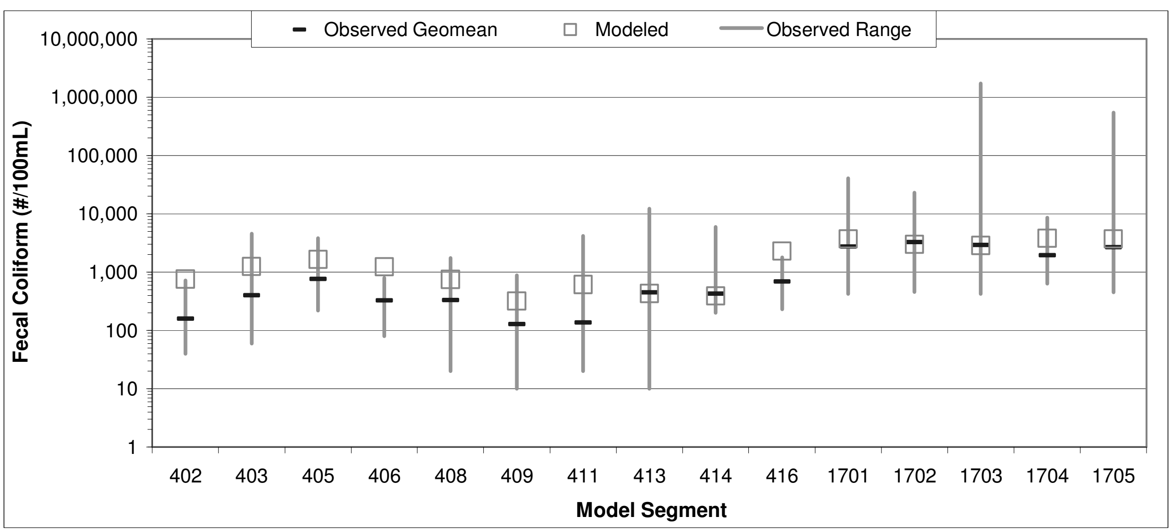


Figure K-13. Validation modeled versus observed fecal coliform concentration for San Juan Creek, Rose Creek and Tecolote Creek (Appendix B, No. 5 and 6)

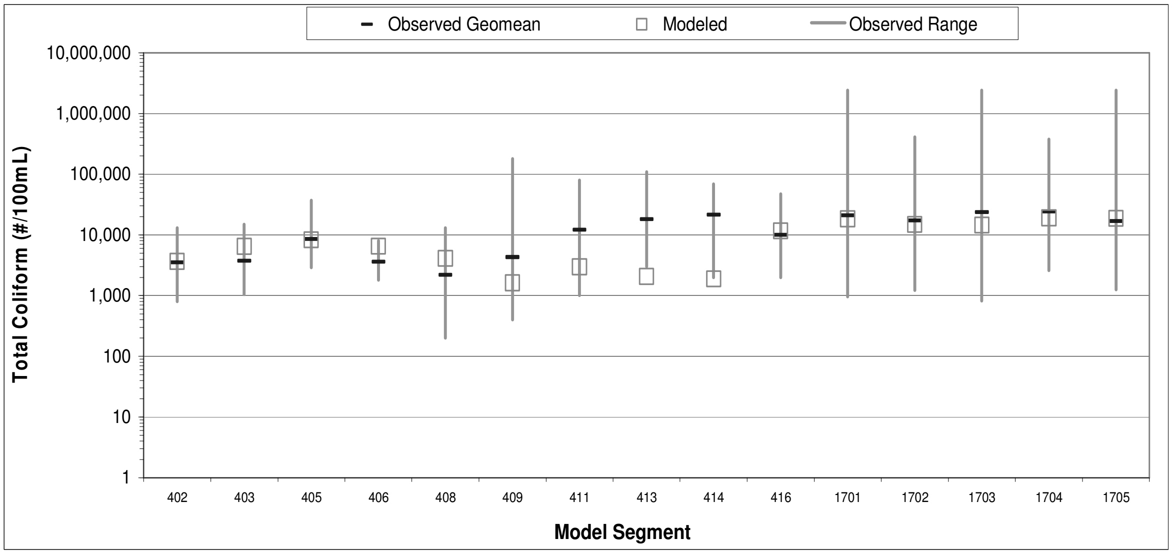


Figure K-14. Validation modeled versus observed total coliform concentration for San Juan Creek, Rose Creek and Tecolote Creek (Appendix B, No. 5 and 6)

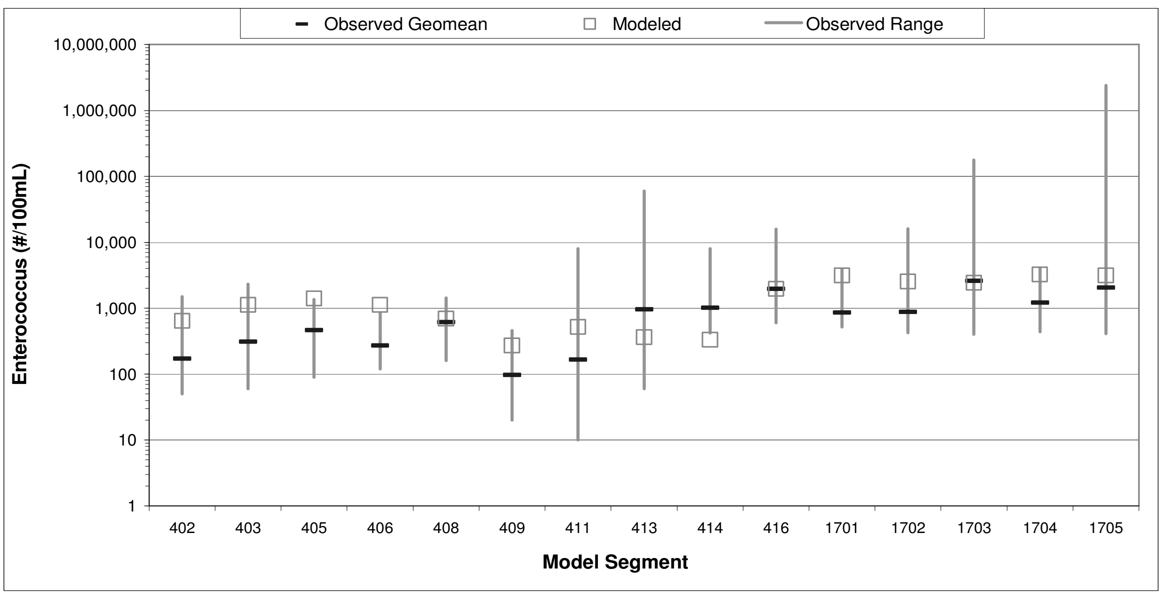


Figure K-15. Validation modeled versus observed enterococci concentration for San Juan Creek, Rose Creek and Tecolote Creek (Appendix B, No. 5 and 6)

.

This page left intentionally blank.